

## **INTEGRATING LEGACY SYSTEMS AND THE GIS ELEMENTS OF THE WATER FRAMEWORK DIRECTIVE INTO A SPATIAL DATA INFRASTRUCTURE**

M. Á. Latre, P. J. Álvarez, R. Béjar, P. R. Muro-Medrano  
latre@unizar.es  
Computer Science and Systems Engineering Department  
University of Zaragoza  
Zaragoza, Spain

### **ABSTRACT**

The Water Framework Directive requires the European Commission to be provided with certain geographic information in the form of maps. The working group in GIS (GIS-GW) was formed to provide common guidance in implementing the GIS elements, presenting a set of best practices that taken to their full extent should led to the creation of an SDI that addresses the GIS requirements of the WFS, according also with the INSPIRE initiative principles and objectives.

These GIS requirements raise challenges in the way the already existing River Basin Authorities have to manage the datasets and information systems they operate with nowadays. This is the case of the Ebro River Basin Authority in Spain. This paper presents how to integrate the organization legacy systems and data into an SDI (focusing on its base: data and its conceptual model) that should allow not only to fulfil the GIS requirements of the WFD, but also to improve the performance of other tasks part of the previous workflow carried out by the organization.

**KEYWORDS:** Spatial Data Infrastructures, Water Framework Directive, Legacy Systems

### **INTRODUCTION**

The Water Framework Directive (WFD), published and entered into force in December 2000, is a legislative framework to protect and improve the quality of all water resources such as rivers, lakes, groundwater, transitional and coastal water within the European Union [European Parliament, 2000]. To achieve a “good status” for all European waters by 2015, a set of activities must be carried out, such as identifying individual river basins and assigning them to river basin districts, determining and characterizing groundwater and surface water bodies, establishing protected areas, setting up water status monitoring networks, reviewing the impact of human activity, analyzing economically the use of water, and ensuring public participation in the establishment of the river basin management plan.

The Water Framework Directive requires the European Commission to be provided with certain geographic information in the form of maps. As part of the Common Implementation Strategy for the WFD [Water Directors, 2001, 2003], a working group in GIS (GIS-GW) was formed to provide common guidance in implementing the GIS elements, presenting a set of best practices that can be optionally followed [Vogt, 2002]. Taken to their full extent, the key points proposed by the GIS-GW should led to the creation of SDIs that address the GIS requirements of the WFD, according also with the INSPIRE initiative principles and objectives [INSPIRE] [JRC, 2002].

This GIS requirements rise challenges in the way the already existing River Basin Authorities have to manage the datasets and information systems they operate with nowadays. This is the case of the Ebro River Basin Authority in Spain [CHE], in charge of physically and administratively managing the hydrographical basin of the Ebro River and also in charge of implementing the WFD elements in their area of influence. This paper presents how to integrate the legacy systems and data of the organization into an SDI that should allow not only to fulfil the GIS requirements of the WFD, but also to improve the performance of other tasks previously carried out by the organization as part of the organization workflow process.

This paper is structured as follows: a brief introduction to the current state of the Ebro River Basin Authority is given. Next, issues related to the integration of the WDF elements and requirements into a legacy system are presented in the next two sections, covering both data level and service and application levels, to finally finish by pointing out some conclusions.

## **THE EBRO RIVER BASIN AUTHORITY**

The Ebro River Basin Authority (Confederación Hidrográfica del Ebro, CHE) is the state organization in charge of physically and administratively managing the hydrographical basin of the Ebro river, the one of the biggest flow in the Iberian Peninsula, through planning (by elaborating and revising a global catchment hydrological plan), managing (by administering and controlling the different water resources in the catchment area) and investing (by projecting and carrying out the public works that may be entrusted to them).

One of the CHE departments which are more involved in the WFD application is the Hydrological Planning Department (OPH). Its administrative work is mostly devoted to the analysis and approval of water point exploitations that are applied by particulars or companies. These exploitations must be granted in conformance with the river basin management plan objectives, and thus this department has to collect and maintain large sets of geodata, which are needed for this process, mainly, a water point inventory that stores over 50,000 water resource points (like wells, springs, drains or reservoir intakes) and the exploitation request files associated to them. This water point inventory it is currently stored in an Oracle 8i spatial database [Béjar et al, 2001] [Latre et al, 2001, 2003]. Workflow process with this inventory includes the gathering of data relative to the water point; analysing the collected data, by comparing it with data about other points, the exploitation data given by the stakeholder and the river basin management plan objectives, to finally deliver several reports about the physic characteristics and location of the point and a resolution about its compatibility with the river basin management plan, allowing or denying the exploitation of the point.

The OPH has also evolved to become a GIS services provider to the rest of the organization, as they are in charge of creating and maintaining most of the hydrogeological datasets they regularly use and frequently create datasets for other departments on request. These data includes not only those which are directly involved in the organization workflow and that is CHE's responsibility and property, which must be collected and maintained; but also those data that are going to be used as a support to daily work. The rest of the information, including data maintained by the CHE, such as water point data, administrative dossier information, superficial and groundwater quality status, hydrogeology and the data obtained from other entities or organizations, like reference geographic data and orthoimages, are stored in a central repository

called *GIS-Ebro* [Arqued et al, 2001], in ArcInfo, shapefiles or raster image formats. Metadata of all of this datasets were created and stored into the database.

The origins of the GIS infrastructure at the OPH consisted of a set of applications oriented to provide tuned functionality for a perfectly established and experienced workflow process, and will be discussed later in this paper. Nowadays, this GIS infrastructure is evolving into an open Spatial Data Infrastructure, which aims at fulfilling the new requirements originated both from the OPH work process, the application of the WFD and the need of interoperation with other organizations.

## **DATA AND METADATA**

### **Conceptual Model**

In order to fulfil the GIS requirements derived from the implantation of the WFD, a new structure and organization for the data was desired. OPH aim is to change from a GIS model based on files (e.g., shapefiles and ArcInfo coverages) into a feature-oriented geodatabase. That is, a conceptual model of the existing data and the data to be collected is needed to, then, analyze the different models of data used by the data sets, match these models with others like the one developed by the GIS-GW to finally define a combined data model and transfer all the existing information into a spatial database [Mustière et al, 2004]. The database management system chosen was Oracle, so the final storage system consisted on integrating into the water point inventory new sets of data. Thus, the database is going to be composed data belonging to the water point inventory, the datasets composing the *GIS-Ebro* repository and the new information to be generated as a result of the WDF works.

The conceptual data model has been built taking into account all this data sources. UML has been chosen as modelling notation, since its becoming an standard methodology, although other notations have been considered, specially MADS, due to its powerful tools to represent spatial feature types, attributes and relations [Parent et al, 1997] [Minout et al, 2004]. The starting point for constructing the conceptual data model was the water point inventory conceptual data model, which was well documented as a result of previous works. The water point inventory model included classes for modelling water points, its hydrophysical characteristics (such as piezometric, hydrometric and chemical time series or lithological column, perforation and lining in the case of wells) and the administrative information associated to them, like water-right owner, water use and water demand.

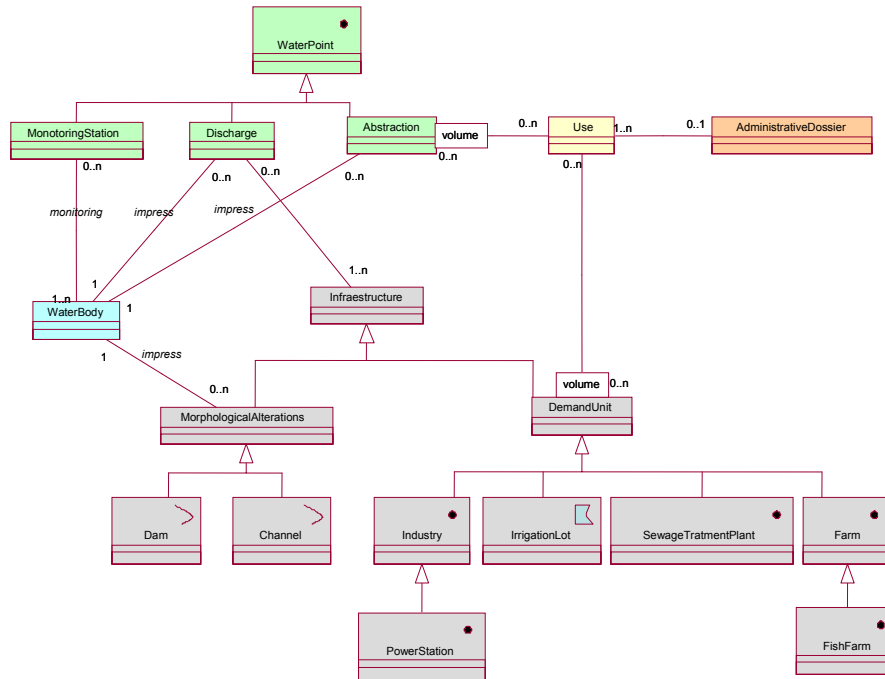
The model was enlarged by including the implicit models used by the ArcInfo coverages and shapefiles of the *GIS-Ebro* repository. Coverages were well documented through metadata, but there was not a conceptual model of them, so relations between different features were missing or, at least, not explicitly set. The data sources were carefully analyzed to determine their subjacent model. The water point inventory model and the *GIS-Ebro* model were tightly integrated into a single model, due to the fact that the water point inventory data were much related to the *GIS-Ebro* data. Relations were even explicit in the water point inventory side, since a set of code tables represented the different features of *GIS-Ebro* with the aim of ensure reference integrity. Obviously, this code tables are no longer used, as it is discussed later.

This water point and *GIS-Ebro* combined model was redefined by adding the feature classes belonging to the WFD domain, and described in the model of the GIS-WG Guidance Document

[Vogt, 2002]. Combining these two models have to be done carefully, analyzing the roles played by the feature classes in order to decide whether to add new classes (like water bodies) or to match two differently named classes [Schwering et al, 2004] (the *MonitoringStation* of the GIS-WG model turned out to be the same as the *ControlPoint* in the water point model). In most cases, however, integrating the WFD classes implied the redefinition of others (splitting of classes or redefinition of relations, e.g., between classes representing rivers and river reaches in the GIS-Ebro and river segments in the WFD model). So, the proposed GIS-GW model is not just put into the global model, but, in fact, a view of the GIS-GW can be made from the global model.

The analysis of pressures and impacts on water bodies requested by the WDF requires the organization to gather and maintain sets of information referring to significant point and diffuse source pollution, water abstraction, water flow regulation and morphological alterations in the case of surface water bodies and artificial recharge in the case of groundwater bodies [European Parliament, 2000]. A subset of the CHE water point database can be viewed, in fact, as a water abstraction database. Furthermore, the administrative process followed to grant or deny water point exploitations provides the OPH with information that can be useful too to detect other pressures and impacts, such as the use the water abstracted is given and the infrastructures on the water bodies that are use to detract water. Data currently stored in the water point database can be used as an starting point to identify water detraction volume, water use and morphological alterations, but a change in the data model has been made to allow the future recollection of this information in a much more structured and useful way (see Figure 1). The new model will allow not only to identify current pressures and impacts, but also to maintain them as part of the regular work process in the OPH. At this point, the OPH staff is considering about integrate completely the other pressures and impacts (point and diffuse source pollution and the rest of anthropogenic impacts) into the model.

The final model was reviewed by incorporating or identifying in the model the classes belonging to the GIS Water Resources Consortium and the ArcGIS Water data model [Davis et al, 2000], in order to support its different points of view (hydrography, drainage, channel features, river network and time series). The result of works derived from the directive will fit in some of these categories. For example, while determining the river water bodies, a drainage network and, next, a river network have been developed (constructed by the Spanish CEDEX in a similar way as described in [Vogt et al, 2003]). The WDF interest stress in the status of the water bodies, but the river and drainage network can be also stored and, thus, contemplated in the model.



**Figure 1.** Partial view of the model linking the administrative work with pressures and impacts

The final model is composed for about 75 classes, and 50 of them are feature classes, that is, have spatial attributes.

### Data storage and migration

Existing data may be adapted to the model, validated and stored in the most convenient way to be used by current information systems and software and to be integrated with the new data collected by the implementation of other tasks of the WFD that should also be incorporated. The storage model elected has been an Oracle 8i spatial database. Procedures for the insertion of new data produced by the other works of the WFD will be needed and defined as these new data are created. In the case of the already existing data, to populate the tables in the database, mappings between the implicit *GIS-Ebro* model and the new model has been developed. This was done by introducing the existing data into temporary tables in the database with the help of the Oracle utilities to be manipulated through SQL queries inside the database afterwards. While performing this process, errors in the original data were detected (mainly due to not fulfilling the desired structure and constraints) and they were corrected following the usual maintenance procedures in the OPH.

Most of the mappings and transformation of the GIS-Ebro data have been able to be performed straightforward, although some, like the case of rivers, have been proved to be more difficult. For instance, there existed two different coverages representing river cartography with different quality and quantity of features. A close analysis of the coverages features revealed that

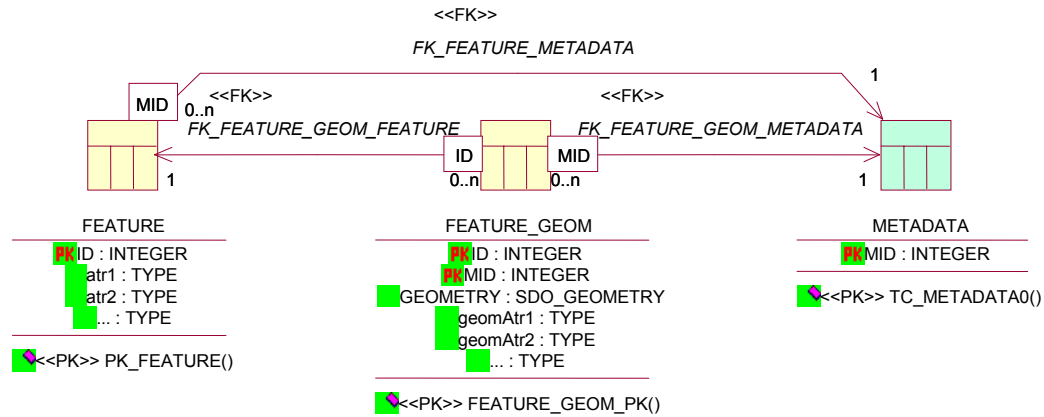
one of them defined a subset of river and river reaches of the other, that is, a subset of the same river network. The main differences were related to the geometry: the less populated coverage had a much more precise geometry than the other. The features were *compatible*, but the geometry was not. The spatial representation pattern we decided to use helped to solve the problem (see Figure 2), providing us with just one river network that can be spatially represented in two different ways. Data from a river shapefile have been mapped into rivers, river reaches and confluences classes. Data migration process lead to the creation of a network in the target model, and next, river confluences had to be identified as nodes to allow the creation and river reaches as edges to allow the construction of a connected network. A preliminary version of the connected network has been done, waiting for the CEDEX (in charge of co-ordinating the surface waters work in Spain) to deliver the final versions of the cartography and the surface water bodies, when the network will be further divided by splitting river reaches that belong to distinct river water bodies into river segments.

Codes and primary keys have also been carefully processed. In the case of existing data (water point inventory and *GIS-Ebro* data), the standards used altogether with alternative identifiers were recognized. The possibility of working with different coding systems has also been contemplated. This is the case of the river coding system. CHE staff is used to work with the rivers' decimal classification (CDR) for both rivers and river reaches. To avoid the disadvantage that it is not possible to add new basins without changing the rest of the river codes, the CEDEX is considering both modifying the CDR system and using the Pfafstetter coding system. The same situation exists with other groups of features. The most appropriate (standard and unlikely to change) code has been chosen and coding-translation tables will be added to the database to allow the use of the other coded systems.

Storage of spatial geometry had also to be treated carefully. The feature classes have been implemented in the database separating the data into two tables, one with the non-spatial attributes and the other with the geometry and the spatial attributes derived from the geometry (see Figure 2). This was done due to several reasons.

In first place, the process is integrating spatial data into a previously existing database that stored mostly non-spatial information. Foreign keys in the IPA database that pointed to tables of codes, were replaced to point to the features stored in the database. After defining the appropriate spatial and non-spatial indexes on the new features, it turned out that non-spatial joins performance was extremely low, due to the geometry field size. Keeping the geometry attribute in a separate table allow the non-spatial join queries to be executed without loss of performance.

In addition to this, the representation features of unknown geometry can be done by the simply absence of a record in the geometry table matching the primary key of the feature, instead of setting to a null value a geometry field. This way, some spatial queries can be defined and executed can be defined written in a simpler way, relaying on the fact that all geometry values in the geometry table are not null.



**Figure 2.** Pattern for storing features into separate tables.  
This schema would belong to a feature class called *FEATURE*

But the main advantage of separating the attributes is that a basic and simple, but effective way to store different representations for the same features is provided. This could be the case of features (like lakes) that can be regarded as polygons when working at a basin scale, but can be useful to be considered as points when working at a European scale. This is also the case of having different precision or quality geometries for the features, like the case of the river previously discussed: all the river features had a geometry (although with position errors), but for the main rivers, a high quality geometry is available in the database. The different geometries should be used alternatively when appropriate. Assuming that different sets of geometry attributes for a certain feature class come from different sources, an identification of the source (MID) can be used to distinguish them. Once the metadata of the information stored in the database are included in the database, this MID will be, indeed, a foreign key pointing to a metadata table describing the source of the geometry.

### Metadata

Metadata of all this data should be generated. There previously existing coverages had their own metadata already created. Mechanisms for create the metadata of the transformed data and the new data produced by the activities of implementing the WFD had to be designed and implemented. The initial intention is, apart from generating metadata for the entire dataset (the database), to generate detailed metadata at feature type level and, when possible and appropriate, at feature instance level.

### SERVICES AND CLIENT APPLICATIONS

The origins of the GIS infrastructure at the OPH consisted of a set of applications oriented to provide tuned functionality for a perfectly established and experienced workflow process. A water point inventory application developed in Java is the core application of the process at the OPH, enforcing the established workflow process. Visualization and spatial data management is done in co-operation with ArcView, customized with some *Avenue* scripts. SIG-Ebro repository information is also available in ArcView, and it is managed with ArcInfo [Latre et al, 2001].

Nowadays, this GIS infrastructure is evolving into an open Spatial Data Infrastructure, which aims at fulfilling the new requirements originated both from the OPH work process, the application of the WFD and the need of interoperation with other organizations. The infrastructure was recently opened to be accessed, on the one hand, by other governmental organizations and subcontracted companies and on the other hand, by general public [Latre et al, 2003]. In the first case, this was done by integrating some standard services compliant with the OpenGIS specification with the existing ones inside the organizations, providing the core of an SDI. Namely, several web map services were added as additional visualization mechanisms to avoid the need of availability of local data. General public access to the data was achieved through a website, where a catalog search tool enables the users to specify queries and obtain information about the list of datasets fulfilling the query restrictions imposed. Users can browse the dataset metadata, access the web map server client to visualize its contents, and, eventually, download the datasets (see Figure 3). Figure 3. Dataset browsing and downloading page (left) and dataset visualization (right)

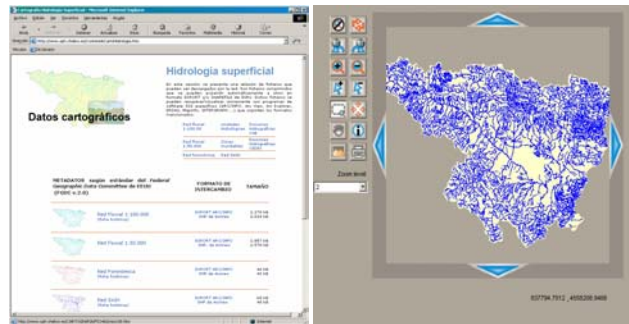


Figure 3. Dataset browsing and downloading page (left) and dataset visualization (right)

Further system functionality improvements are going to be directed to expand SDI capabilities of the system, limiting the ad-hoc software development and trying to define and implement standard interoperable web services and increasing the opportunities of reuse of these services.

Firstly, standard services are planned to be used in order to detach the system from commercial applications. Major modifications of the water point inventory application to be done (management of water demand units) involve managing new types of elements with geometries different from point geometry. This is planned to be achieved by the use of web applications accessing standard OpenGIS web services instead of communicating and co-ordinating this task with ArcView. Other functionality in the legacy system that can be replaced by the standard services has been identified. A gradual change directed to rely on standard services will lead to an increment of the SDI capabilities and a reduction of the ad-hoc, non-standard elements, decreasing maintenance efforts

Secondly, it is needed to provide access to the WFD related data in the terms proposed by the GIS-GW in the short and long term. OpenGIS standard web map services will be reused to provide long term data access mechanism proposed. For the short term, web feature services will be created to support the short term data access mechanism in GML format, and it will serve as a base for providing the data in a shape file format, not only as a minimum data exchange mechanism between Member States and the European Commission, but also inside the organization or as a

facility to the general public. The use of OpenGIS standard services can be a way to achieve syntactic interoperability among the services run by other authorities, whilst the use of a data model compatible with the one proposed by the GIS-GW should ensure semantic interoperability.

Finally, some non-hydrogeological classes are also integrated into the model (like, for instance, administrative limits). For those feature types that are not under the responsibility of the CHE and, consequently, are not maintained by them, separate standard services are planned to be developed. The goal is to be able to replace them by their equivalent (and compatible) services provided by the authoritative organizations (in the case of administrative limits, they should be the web services of the Spanish SDI, in charge of the IGN). Once that strategy is a reality, services created inside the CHE are planned to be maintained as a backup, to prevent the organization activity of being stopped by any hardware or software failure outside the organization.

## CONCLUSIONS

The GIS requirements of the WFD rise challenges in the way the already existing River Basin Authorities have to manage the datasets and information systems they operate with nowadays. This is the case of the Ebro River Basin Authority in Spain. This paper has presented how the integration of the organization legacy systems and data along with the requirements of the WFD is being performed in this organization.

The main works to meet the requests of the WFD have not been contracted so far, so we have presented the preliminary works that have been carried out, mainly at the data level, and how we pretend the future GIS works to be. The ideas presented here are also supported by the OPH and the CHE staff, although final results will probably depend on the methodology of the firm or groups of firms the work will be finally assigned to.

## ACKNOWLEDGEMENTS

The basic technology of this work has been partially supported by the Spanish Ministry of Science and Technology through the project TIC2003-09365-C02-01 from the National Plan for Scientific Research, Development and Technology Innovation.

## BIBLIOGRAPHY

- Arqued Esquía, V. M., Losada García, J. A., Zarazaga Soria, F. J. (2001). El sistema de información GIS-Ebro. Metadatos y catálogo de datos geográficos. *Bole.tic, septiembre – octubre 2001*.
- Béjar, R., Fernández, P., Latre, M.Á., Muro-Medrano, P.R., Rioja, R., Zarazaga, J. (2001). Interoperation of distributed GIS components in a Hydrological Data Inventory System. *4<sup>th</sup> AGILE Conference: GI in Europe: Integrative – Interoperable – Interactive*.
- CHE, Confederación Hidrográfica del Ebro (Ebro River Basin Authority), <http://www.chebro.es>.
- Davis, K., Whiteaker, T., and Maidment, R. (2000). Definition of the Arc Hydrology Data Model. *GIS in Water Resources Conference, University of Texas at Austin 23-25 February, 2000*. <http://www.crrw.utexas.edu/giswr/resources/library/archydro.pdf>.
- European Parliament and Council of the European Union. (2000). Directive 2000/60/EC (Water Framework Directive), *Official Journal of the European Communities L 327*, [http://europa.eu.int/comm/environment/water/water-framework/index\\_en.html](http://europa.eu.int/comm/environment/water/water-framework/index_en.html).

- INSPIRE. The Infrastructure for Spatial Information in Europe (INSPIRE) initiative homepage, <http://www.ec-gis.org/inspire>.
- JRC-Institute for Environment and Sustainability. (2002). INSPIRE Architecture and Standards Position Paper, [http://inspire.jrc.it/reports/position\\_papers/inspire\\_ast\\_pp\\_v4\\_2\\_en.pdf](http://inspire.jrc.it/reports/position_papers/inspire_ast_pp_v4_2_en.pdf).
- Latre, M. Á., Béjar, R., Muro-Medrano, P. R. (2003). Distributed Access to a Hydrological Administrative Spatial Data Infrastructure, *Proceedings, 9<sup>th</sup> EC-GI&GIS Workshop ESDI, June 25-27 2003, A Coruña, Spain*.
- Latre, M. Á., Fernández, P., Cantán, Ó., Garcés, C., Rioja, R., Molinero, J. (2001). Interoperabilidad de componentes SIG en un Sistema de Información Hidrológico, *JSIG 2001*.
- Minout, M., Parent, C., Zimányi, E. (2004). A Tool for Transforming Conceptual Schemas of Spatio-Temporal Databases with Multiple Representations, *Proceedings of the IASTED International Conference on Database Applications, DBA'2004, Innsbruck, Austria*.
- Mustière, S., Cleach, A. and Fort, J. (2004). Unifying Databases: Experiments with a Hydrographic Database, *Proceedings, 7th AGILE Conference on Geographic Information Science, Heraklion, Greece*.
- Parent, C., Spaccapietra, S., Zimányi, E., Donini, P., Plazonet, C., Vangenot, C., Rognon, N., Pouliot, J., Crausaz, P. (1997). MADS, modèle conceptuel spatio-temporel, *Revue Internationale de Géomatique*, 7(3-4): 317-352.
- Schwering, A., Hart, G. (2004). A Case Study for Semantic Translation of the Water Framework Directive and a Topographic Database, *Proceedings, 7th AGILE Conference on Geographic Information Science, Heraklion, Greece*.
- Vogt, J. (ed.). (2002). Guidance Document of Implementing the GIS Elements of the Water Framework Directive, *European Communities*  
[http://forum.europa.eu.int/Public/irc/env/wfd/library?l=/framework\\_directive/guidance\\_documents/guidance\\_outcome/wg\\_gis\\_guidance/EN\\_1.0\\_&a=d](http://forum.europa.eu.int/Public/irc/env/wfd/library?l=/framework_directive/guidance_documents/guidance_outcome/wg_gis_guidance/EN_1.0_&a=d).
- Vogt, J., Colombo, R., Paracchini M. L., Jager, A., Soille, P. (2003). CMM River and Catchment Database, *EUR 20756 EN, EC-JRC, Inspira*.
- Water Directors (2001). Common Implementation Strategy for the Water Framework Directive. Strategic Document. <http://europa.eu.int/comm/environment/water/water-framework/strategy.pdf>
- Water Directors (2003). Carrying forward the Common Implementation Strategy for the Water Framework Directive. Progress and Work Programme for 2003 and 2004.  
<http://europa.eu.int/comm/environment/water/water-framework/strategy2.pdf>